SOCIETY OF American Archivists

Technical Subcommittee on Encoded Archival Standards

Quo vadis, EAD?

Current use cases - from single institutions to aggregators and providers of archival management systems

2021-06-04

Presenters
Karin Bredenberg, co-chair TS EAS
Lindsey Loeper, University of Maryland, Baltimore County
Ada Negraru, Southern Methodist University, Dallas
Greg Wiedeman, University at Albany, The State University of New York

CC BY NC SA

TS-EAS

- Technical Subcommittee on Encoded Archival Standards at the Society of American Archivists (SAA)
- Background and work covered in this presentation:
  - https://youtu.be/9NXNyx9py-I

Where you can find more

TS EAS
https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas

TS-EAS on GitHub
https://github.com/SAA-SDT

EAD publication
http://www.loc.gov/ead/index.html

EAC publication
https://eac.staatsbibliothek-berlin.de/

Our mailing list
EAD@LISTSERV.LOC.GOV

Reporting an issue via SAA
https://www2.archivists.org/standards/TS-EAS-report-an-issue

Standards revision

Annual rolling revision cycle for minor releases
(see more on GitHub: https://github.com/SAA-SDT/TS-EAS-subteam-notes/blob/master/rolling-revision-cycle.md)

Evaluate standards for potential major revision every five years following guidelines by SAA's Standards Committee
(see more: https://www2.archivists.org/governance/handbook/section7/groups/Standards/Development-and-Review)

Presentations

Repurposing descriptive data:

PastPerfect, EAD3, & PHP

Lindsey Loeper
Reference & Instruction Archivist
Special Collections, Library
University of Maryland, Baltimore County
@lindseyloeper

bit.ly/umbcead

UMBC

SAA TS-EAD, June 4, 2021

Pro: Lots of access points

Collection level record in PastPerfect and PastPerfect Online
Collection level record in MARC - University catalog and Worldcat
Multi level finding aid in Word and PDF
Multi level finding aid in PHP

Con: Moving data between these access points

Limited extraction and reformatting options
Almost all description is done by hand - copy, paste, copy, paste, etc.
Lots of room for mistakes and requires very careful QC at each step

NHPRC Access to Collections grant, 2014-2016
"Implementing an EAD-XML compliant workflow at the UMBC"

Goals for the new workflow

- ➔ Collection level records for all archival collections
- ➔ Repurpose data across platforms
- ➔ Include creation of MARC records
- ➔ Minimize potential for future backlogs
- ➔ Make our work accessible and adaptable
  - ◆ GitHub: /UMBC-Library/EAD-XML
  - ◆ bit.ly/umbcead

Data guidelines and clean-up

Arrangement

PastPerfect field	Guidelines/Rules	Description level
System of Arrangement <System>	DACS: 3.2 System of Arrangement (Added Value)	Collection - Required
	All processed collections must have this field completed to the Series level (do not include sub-series). Include series number, series title, and inclusive dates. Unprocessed or accession level records should document any arrangement present when accessioned and/or if any arrangement work was completed when accessioned. May be used for item level records if applicable. Format as a paragraph, not a list. May also be expanded in the finding aid to include sub-series, or to display as a list, if desired. Assign series with Roman numerals and subseries with capital letters. Example: Series I. Series I.A. Series I.B. Series II. Series II.A.	Accession - Required
PastPerfect screen location		Item - Optional
Archive > Structure		Include in export?
		PastPerfect Online, EAD MARC
Description		MARC field
Outlines how the collection or materials being described are currently organized.		351 $a
Example	EAD tag	
Organized into three series: Series I. Personal, 1839-1994 Series II. Sun Papers, 1927-1995. Series III. Boordy Vineyard, 1933-1995.	<archdesc><arrangement encinganalog="351$a"> Organized into three series: Series I. Personal, 839-1994. Series II. Sun Papers, 1927-1933-1995. </arrangement>	

PastPerfect export to PastPerfect XML

PastPerfect XML transformed using XSLT:

- EAD-XML (EAD3)
- MARC XML

EAD-XML transformed using in-house batch processor to create PHP for online finding aid.

```
<date>1839-1995; bulk 1942-1989</date>
<earlydate>1839</earlydate>
<latedate>1995</latedate>
```

```
<unitdate unitdatetype="inclusive">1839-1995</unitdate>
<unitdate unitdatetype="bulk">bulk 1942-1989</unitdate>
<unitdatestructured unitdatetype="inclusive">
    <daterange>
        <fromdate>1839</fromdate>
        <todate>1995</todate>
    </daterange>
</unitdatestructured>
```

```
<table>
    <tr><td>Dates:</td>
    <td>1839-1995; bulk 1942-1989</td></tr>
</table>
```

XSLT

```
<!-- Separation of Bulk and Inclusive Dates -->
<xsl:choose>
    <!-- If bulk date delimiter (;) exists -->
    <xsl:when test="contains(date,';')">
        <unitdate unitdatetype="inclusive" encodinganalog="245$f">
            <xsl:value-of select="normalize-space(substring-before(date,
';'))"/>
        </unitdate>
        <unitdate unitdatetype="bulk" encodinganalog="245$g">
            <xsl:value-of select="normalize-space(substring-after(date, ';'))"/>
        </unitdate>
    </xsl:when>
<!-- If only inclusive dates exist -->
    <xsl:otherwise>
        <unitdate unitdatetype="inclusive" encodinganalog="245$f">
            <xsl:value-of select="date"/>
        </unitdate>
    </xsl:otherwise>
</xsl:choose>
```

Room for improvement: Subjects

<subjects>Universities and colleges -- History;
Universities and colleges -- Planning;
Schamp, H. W. (Homer Ward), 1923-;
Dorsey, John W., 1936-;
University of Maryland, Baltimore County;
Lee, Calvin B. T.;**</subjects>**

Thesauri information and XML tags not exported from PastPerfect. Requires hands-on clean up post export and careful review by the Catalog Librarian.

Single level / collection level records

1. PastPerfect Online (+ search engines)
2. MARC records in USM ILS
3. MARC records in UMBC discovery layer
4. MARC records in WorldCat
5. MARC records in ArchiveGrid

Multi level records / finding aids

1. EAD-XML files
2. EAD-XML files in ArchiveGrid
3. EAD-XML files in NLM consortium
4. PHP files (+ search engines)
5. PHP files linked from all single level records
6. PDF files for printing

For more information…

GitHub: /UMBC-Library/EAD-XML
bit.ly/umbcead

Lindsey Loeper
lindseyloeper@umbc.edu
@lindseyloeper

THE TRUSTED OLD FRIEND: EAD 2002 and Special Collections Finding Aids

ADA NEGRARU, DEGOLYER LIBRARY, SOUTHERN METHODIST UNIVERSITY

QUO VADIS, EAD? SAA TS-EAS, JUNE 4, 2021

SOUTHERN METHODIST UNIVERSITY
Special Collections Libraries

- Southern Methodist University (SMU): medium-size private university in Dallas, Texas
- Special Collections within the SMU Libraries: DeGolyer, Bridwell, Hamon (Bywaters) and the University Archives.
- Types of materials: rare books, manuscripts, photographs, university archives, corporate records, artwork, mixed media, video and audio recordings, and others.
- Discoverability
  - MARC bibliographic records
  - EAD finding aids
  - Digital collections (digitized materials)

EAD 2002 Finding Aids

- EAD2002 adopted in the early 2000s; still in use
- Conforms to DACS
- Well suited for archival and research purposes (description and discoverability)
- Published by Texas Archival Resources Online (TARO) and harvested by ArchiveGrid.
- 370 finding aids published to date

HOW DOES IT WORK?

- **Collaboration**: Campus- wide EAD working group
- **Communication**: shared network drive, email, group meetings
- **Responsibilities**: finding aid creation; authority terms; revision; coordination
- **Workflow:** finding aid creation to publication

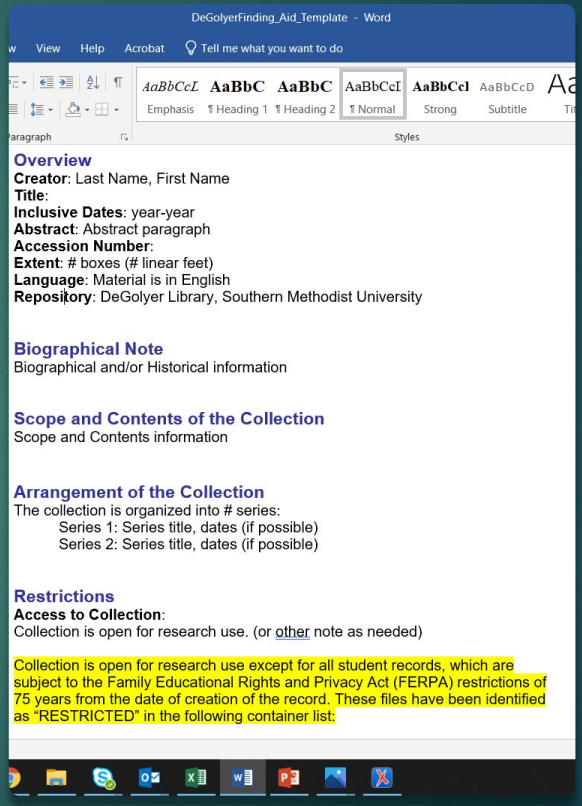
MORE ABOUT WORKFLOW

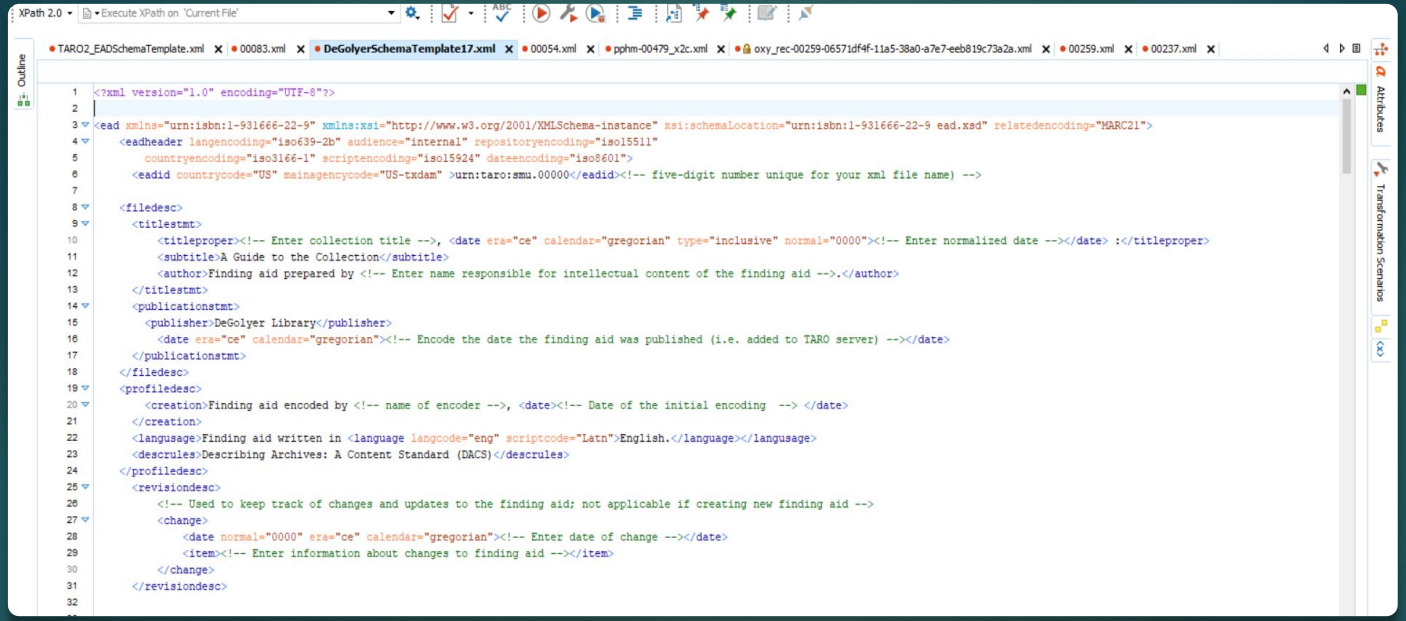
IT'S ALL IN THE TEMPLATE

- EAD 2002 used to create finding aid templates
  - Text documents : Word
  - EAD-DTD schema - XML template: Oxygen
  - Integrates well with state consortium: CSS stylesheets
  - All needed descriptive elements: Title, Creator, Dates, Identification, Language(s), Scope and Contents, index terms, other repository and collection specific notes

Finding aid template – Word

- Used by archivists, interns, student assistants
- Simple training process
- Enough depth for collection description and research
- Flexibility – can accommodate new institution policies (e.g. Sensitive materials note)





EAD2002 – DTD schema template

- Software: Oxygen Editor for Academic Use

EAD 2002 – DTD schema

- Created following consortium (TARO) requirements
- Adapted to institution needs
- Straightforward encoding
- In-house training process

WHAT ABOUT THE FUTURE?

- **EAD 2002 vs. EAD3**
  - EAD3 has been considered and evaluated
  - EAD2002 serves our purposes short to medium term

- **Why not EAD3 right now**
  - Too structured for our current needs
  - Impact on current templates and workflow
  - Impact on staff workload; collection processing, other tasks would be affected
  - Training costs
- **Potential reasons to adopt EAD3**
  - Linked data
  - User expectations
  - If required for continued participation in TARO, ArchiveGrid, NAFAN

From EAD to a Common Data Model

gregwiedeman.com/slides/EAD2021.html

Gregory Wiedeman
University Archivist
University at Albany, SUNY

My experience with EAD

- Learned technology by hand-encoding EAD
- Built tools to work with EAD
- Work with ASpace, ArcLight, Hyrax/Samvera
- Manage traditional archival processes
- EmpireADC advisor

The Promise of EAD

"An encoding standard would guarantee that the machine-readable finding aids created today would be usable tomorrow!"

Daniel V. Pitti, "The Berkeley Finding Aid Project," 1994
http://archive1.village.virginia.edu/dvp4c/arlpap.htm

One Final Format

- Structure your archival description
- Preservation format, future-proof
- Easily move data between systems

One Final Format

- Structure your archival description
- Preservation format, future-proof
- Easily move data between systems

We need to give up on this dream

Archival systems still limited

- ArchivesSpace and other successful tools
- Front-end usability is still a problem
- No archival system to manage digital objects
- Funding and sustainability are challenges

Archival systems still limited

- ArchivesSpace and other successful tools
- Front-end usability is still a problem
- No archival system to manage digital objects
- Funding and sustainability are challenges

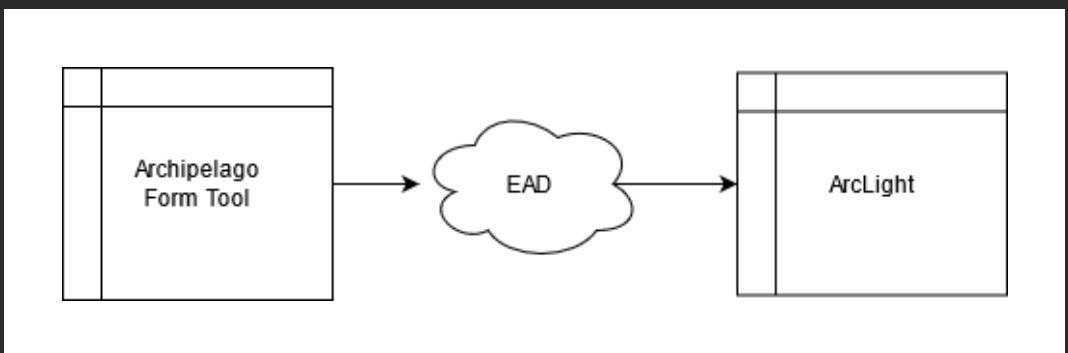
As the data structure standard, EAD defines how systems work with archival description

Hard to build systems with EAD

- Variability
  - `<unitdate>` within `<unittitle>`
- Mixed content
  - `<languagenote>`, `<title>`, `<date>`, etc.
- Designed with document in mind
- EAD3 is incremental improvement, adds more variability
  - `<languagedeclaration>`
  - `<unitdatestructured>`

EAD is a Lossy Medium

- EAD does not have strict rules
- "Migration" problems
- EmpireADC example:

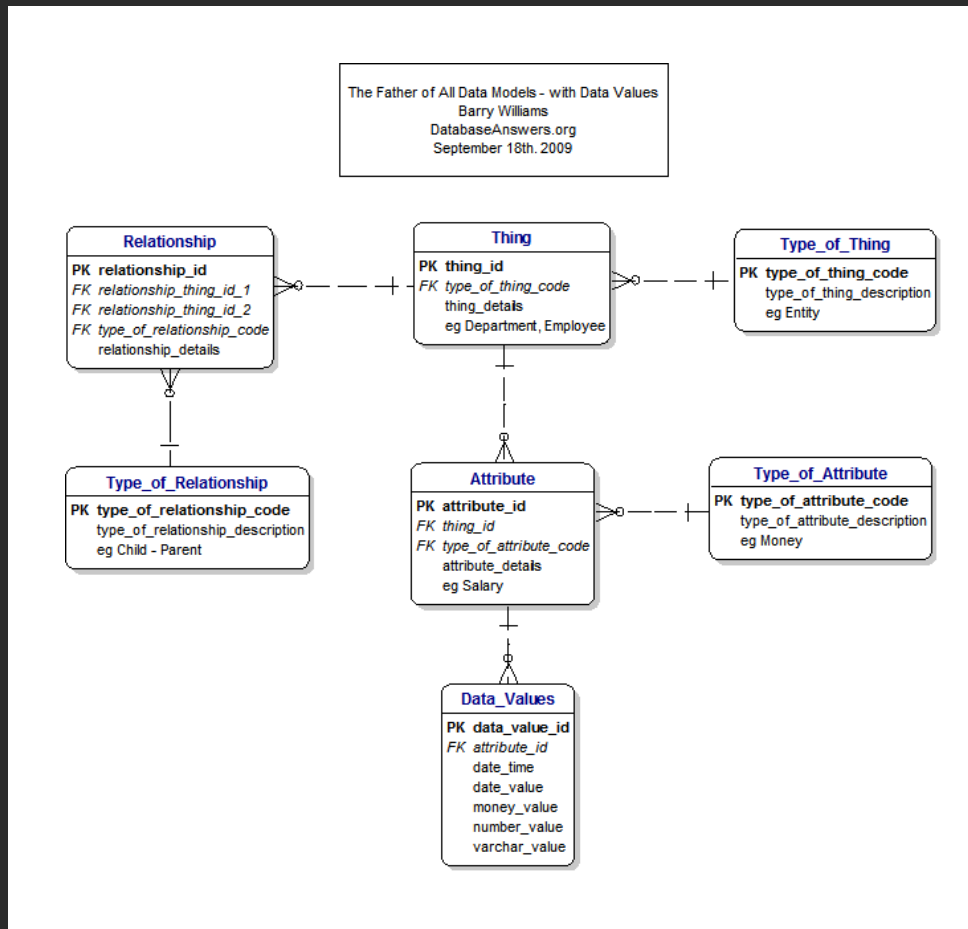


EAD is Limiting for Aggregators

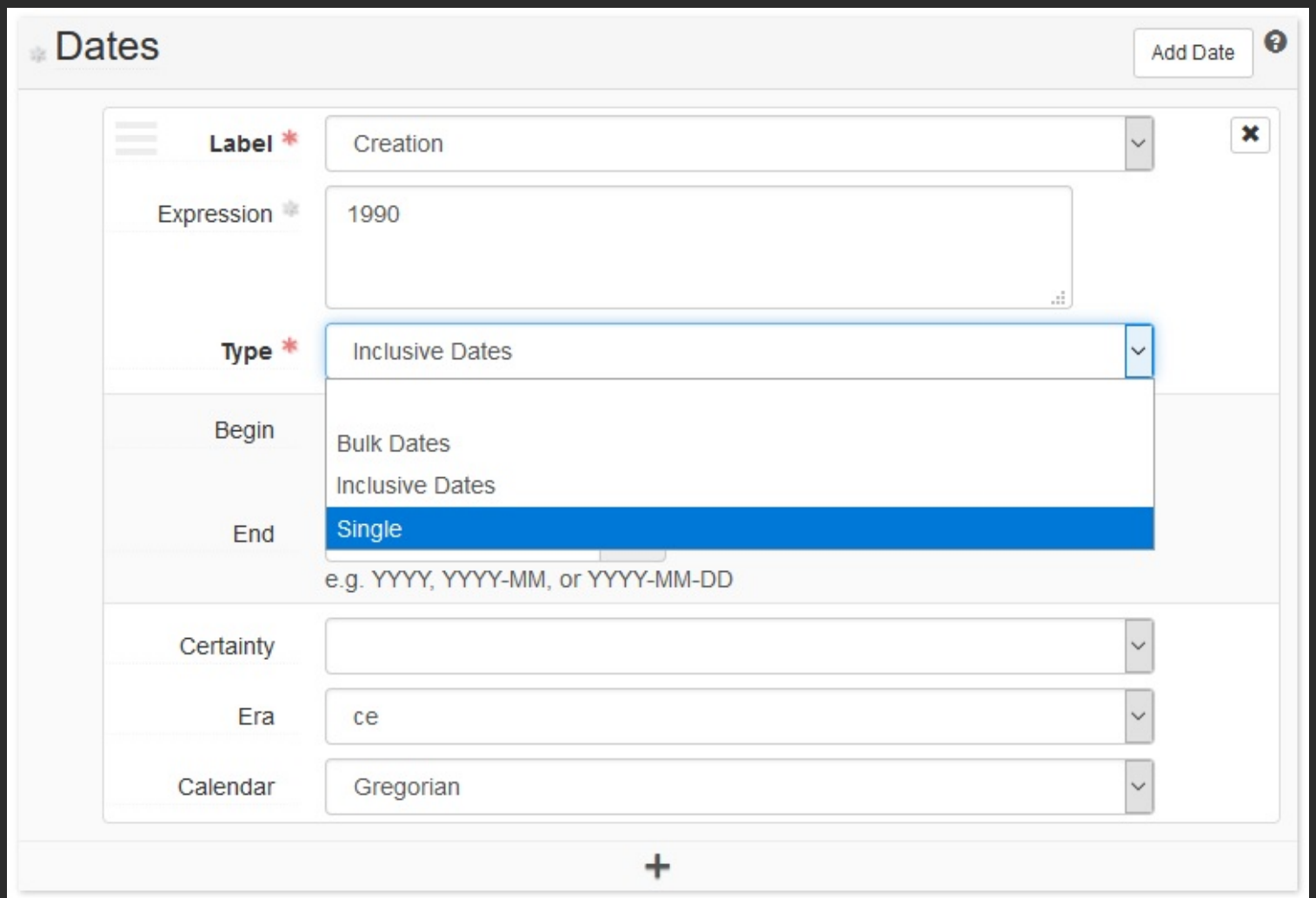
- "Raw" EAD is challenging to index
  - Hand-encoded
  - Custom authoring tools
- Requires skilled labor to normalize
- Aggregators need to rely on consistent EAD
  - ArchivesSpace
  - EmpireADC using Archipelago-based "form tool"

From XML format to Data Model

- Accept that description will live in multiple formats and systems
- Standardize data structure, not format



Date Problems

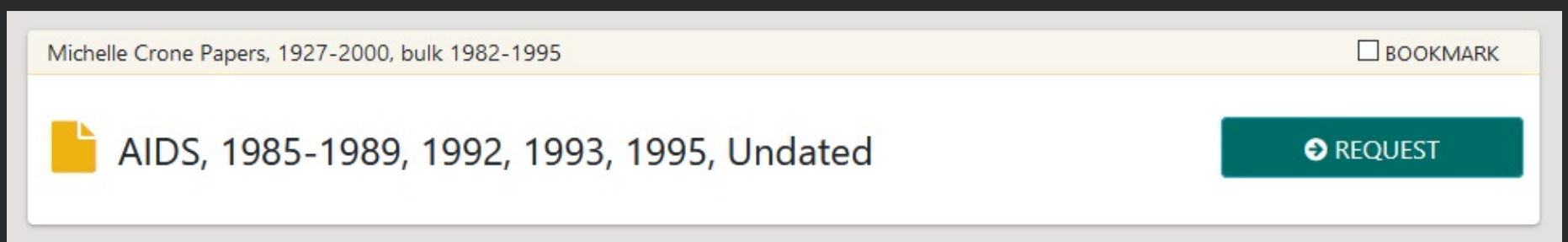


- EAD2002:
  - TYPE – not required. May be used to specify whether the date is “bulk” or “inclusive.”
- EAD3:
  - @unitdatetype – not required. Use to record whether the unit dates are bulk or inclusive. Possible values: bulk, inclusive.

ASpace exported "single" `<unitdate>` without `@type`

Date types in ArcLight

- Combines dates to Solr field: `normalized_title_ssm`



ArcLight Bug

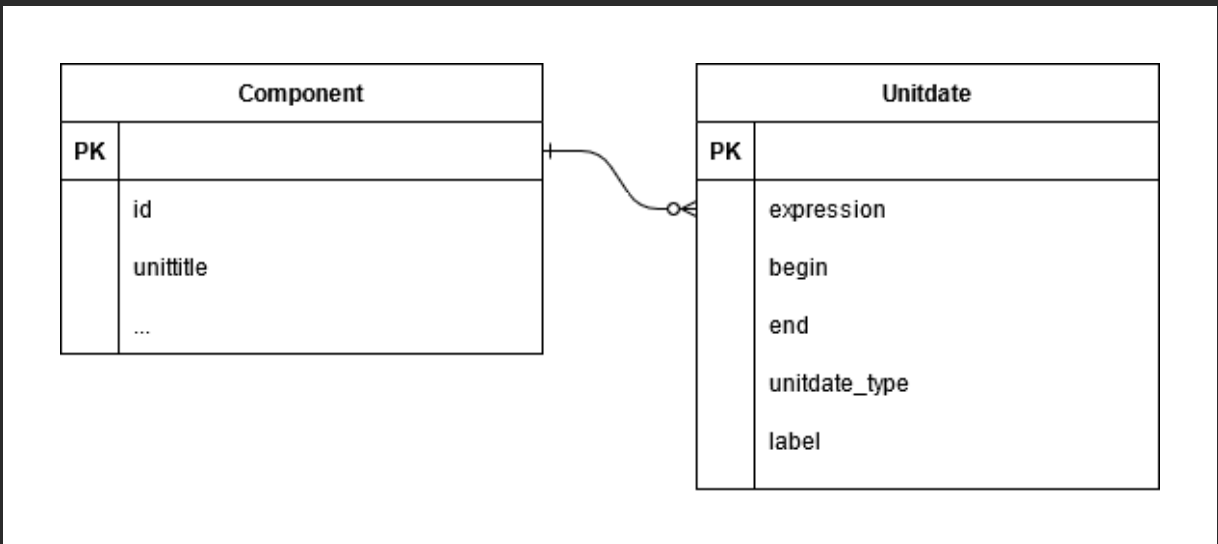
- Originally assumed one date
- then expected `@type="inclusive"` or `@type="bulk"`

General Administrative Files, 1965-1966, 1978-1990, 1963, 1994, undated

Date Problems

- Misunderstanding between tools
- EAD does not clearly define the structure of dates

Data Model: A Set of Rules



Expressed in JSON

```
{
  "expression": "1958 May 31-1960 June 2",
  "begin": "1958-05-31",
  "end": "1960-06-02",
  "date_type": "inclusive",
  "label": "creation"
}
```

Expressed in EAD2002

```
<unitdate label="creation" type="inclusive" normal="1958-05-31/19
    1958 May 31-1960 June 2
</unitdate>
```

Expressed in EAD3

```
<unitdatestructured label="creation" unitdatetype="inclusive">
    <dateset>
        <daterange>
            <fromdate standarddate="1958-05-31">1958 May 31</from
            <todate standarddate="1960-06-02">1960 June 2</todate
        </daterange>
    </dateset>
</unitdatestructured>
```

Serialization is easier

- Let the technologists deal with that
- Rely on common conventions for implementations
  - EAD2002
  - EAD3
  - JSON
  - Records in Context ontology (RiC-O)
  - Spreadsheets

EAD already relies on conventions

```
<extent>32 cubic ft.</extent>
<extent units="cubic ft.">32</extent>
```

More inclusive approach

- Migtation will be required at some point
- Lowers cost of systems and integration
- Allows new possibilities for small repositories
  - Google Forms
  - Spreadsheets
- Allows broader professional participation

From EAD to a Common Data Model

gregwiedeman.com/slides/EAD2021.html

Gregory Wiedeman
University Archivist
University at Albany, SUNY

Thanks!

- The recording will be available on YouTube:

  https://www.youtube.com/user/saastaff

- Feedback? Suggestions for ...? Please complete our survey!

  http://bit.ly/eadwebinar